

SIG e regressão logística para mapeamento de risco de contaminação por pesticidas nos mananciais superficiais da bacia do Alto Rio das Mortes-MT

Peter Zeilhofer¹
Ivani Matos de Oliveira²
Suzy Mara Klemp¹
Emerson Soares dos Santos¹
Eliana Freire Gaspar de Carvalho Dores³

¹ Universidade Federal de Mato Grosso (UFMT), Departamento de Geografia
Av. F. Corrêa da Costa, s/n - 78060-900 - Cuiabá - MT, Brasil
pitalike@terra.com.br / suzy_klemp@yahoo.com.br / emer.santos@uol.com.br

² Universidade Federal de Mato Grosso (UFMT), Mestranda em Física e Meio Ambiente
Av. F. Corrêa da Costa, s/n - 78060-900 - Cuiabá - MT, Brasil
ivanigeo@ufmt.br

³ Universidade Federal de Mato Grosso (UFMT), Departamento de Química
Av. F. Corrêa da Costa, s/n - 78060-900 - Cuiabá - MT, Brasil
eliana@cpd.ufmt.br

Abstract. Present study evaluates the applicability of GIS and Logistic regression analysis for contamination risk mapping by pesticides in streams of the Rio das Mortes watershed, one of the most productive crop farming region in the Mato Grosso state. Water quality data was sampled at five sub-watershed outlets, totalizing 109 campaigns. A SRTM digital elevation model was processed for watershed extraction and determination of slope. A distance map was derived from digitalized stream network. Thematic layers on soil types, geomorphologic units and agricultural suitability were obtained from state socio-environmental zoning. Landsat ETM and CBERS CCD imagery were visually interpreted for land cover mapping. Influence of watershed characteristics on pesticide presence were evaluated by Logistic Regression (RL) models. Final RL model, which identified season and the distance layer as significant explanatory variables, was then applied to spatially predict potential pesticide contamination risk of surface water.

Palavras-chave: GIS, water quality, pesticide contamination risk, logistic regression, Sistemas de Informação Geográfica, qualidade de água, risco de contaminação por pesticidas, regressão logística.

1. Introdução

As relações causais entre uso agrícola na qualidade de água foram comprovadas em diversos estudos, entre eles os de Fisher *et al.* (1999) e Stow *et al.* (2001). Em países com intensa atividade agrícola tais como os Estados Unidos, a poluição difusa ou não-pontual é considerada o maior problema para a qualidade de água (EPA 2003), introduzindo nos ecossistemas aquáticos sobrecargas de nutrientes, pesticidas e sedimentos. Como base de estudos sobre impactos e origens de problemas de qualidade de água, técnicas de sensoriamento remoto e geoprocessamento podem ser aplicados no mapeamento de uso e ocupação de solo e na parametrização das bacias de contribuição. Basnyat *et al.* (2000) e Tong & Chen (2002) utilizaram imagens dos sensores Landsat TM e SPOT em conjunto com técnicas da análise espacial para estudos de problemas de poluição não-pontual. Munfano *et al.* (2006) desenvolveram um índice de potencial de poluição difusa, chamado *PNPI*, baseado em planos de informação derivados de um modelo numérico de terreno (MNT), mapa de uso e ocupação de solo e mapa geológico. Na região Cento-Oeste de Brasil, Zeilhofer *et al.* (2006) aplicaram técnicas de geoprocessamento tais como análise de MNTs, classificação de imagens Landsat ETM e buscas espaciais para estudos dos padrões geográficas da qualidade das águas superficiais na Bacia do Rio Cuiabá.

Diversas técnicas uni- e multivariadas foram propostas para estudos das relações entre características sócio-ambientais de bacias, extraídas por técnicas de geoprocessamento e parâmetros de qualidade de água. Devido à característica multidimensional de dados físico-químicos de qualidade de água, técnicas de ordenação, tais como análise de componentes principais (ACP), possuem vasta aplicação (Wunderlin *et al.* 2001, Bengraïne & Marhaba 2003). Outra técnica multivariada, a análise de covariância (ANCOVA) foi utilizada por Donohue *et al.* (2005) para avaliação dos padrões espaço-temporais de frações de nitrogênio e fósforo em mananciais superficiais da bacia do rio Lough Mask, Irlanda.

Em modelações espaciais, Regressões Logísticas (RL) são utilizadas quando a variável resposta é dicotômica, para estimativas da probabilidade de um evento (ex.: presença de um poluente, concentração de um poluente acima de uma determinada concentração). RLs foram amplamente utilizadas para modelação espacial de risco de poluição de águas sub-terrâneas (Green *et al.* 2004). Estudos de casos para avaliação da ausência/presença de agrotóxicos em mananciais superficiais foram apresentados por Battaglin *et al.* (1998) e Battaglin *et al.* (2003).

Os planaltos sedimentares do Centro Oeste experimentaram nas últimas três décadas uma transformação intensa das suas formações vegetais naturais de Cerrado em áreas agrícolas, principalmente para cultivo de soja e algodão. Os conhecimentos sobre os impactos desta atividade agrícola, entretanto, são limitados ainda. Neste contexto, presente estudo avalia a aplicabilidade conjunta de Sistemas de Informação Geográfica e técnicas estatísticas para mapeamento de risco de contaminação por pesticidas nas áreas de contribuição do Alto Rio das Mortes.

2. Área de estudo

A bacia do Alto Rio das Mortes é uma das mais produtivas áreas de plantio de soja e algodão no estado de Mato Grosso. Possui temperaturas médias mensais entre 23 e 27 °C e precipitação média anual de cerca de 1660 mm (1981 a 1998) (Dores & De-Lamonica-Freire 2001), com concentração de cerca de 80% das precipitações entre os meses de Novembro até Abril.

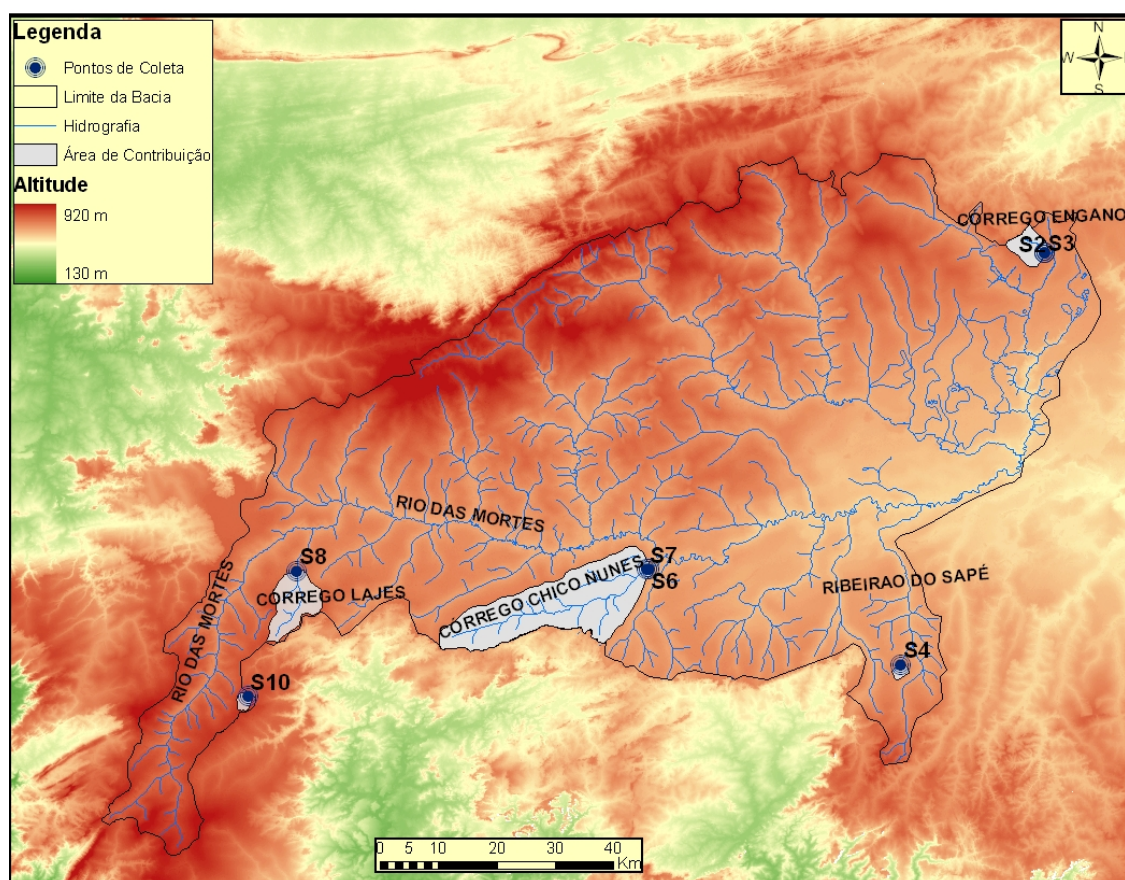


Figura 1. Rede hidrográfica e MNT da bacia do Alto Rio das Mortes, sobreposto com as sub-bacias amostradas.

3. Materiais e Métodos

Planos de informações (PIs) temáticos da rede hidrográfica, das unidades litológicas, tipos de solo e aptidão agrícola na escala 1:250.000 em formato vetorial foram obtidos junto a SEPLAN (2004), provenientes do zoneamento agro-econômico-ecológico do estado de Mato Grosso. Foi calculado, a partir de uma função “buffer” camada matricial contendo a distância de cada célula do manancial mais próximo, como representado no PI da hidrografia, escala 1:250.000. Um modelo numérico de terreno (MNT) da missão SRTM com resolução horizontal de 90 m foi obtido junto a EMBRAPA, do qual foram derivados, a partir de procedimentos padrão de análise espacial, os limites das sub-bacias e PI matricial da declividade. Imagens multitemporais dos sistemas sensores CBERS CCD do ano 2006 (cenas 165/117, 164/117, datas 22/06/2006 e 25/06/2006) e Landsat ETM dos anos 2002 e 2003 (225/70, 225/71, 226/70, 226/71) foram interpretadas visualmente para mapeamento de uso e ocupação do solo durante o período das coletas de qualidade de água. A interpretação foi subsidiada por imagens de alta resolução disponível pelo sistema “Google Earth”.

Todos planos de informação foram georeferenciados para a projeção UTM (SAD 69), utilizando as bases cartográficas 1:250.000 da SEPLAN (2004). Os trabalhos de geoprocessamento foram efetuados a partir dos softwares SPRING 4.2 (INPE, São José dos Campos) e ArcGIS 9.1 (ESRI, Redlands).

Amostras de qualidade de água foram coletadas em cinco sub-bacias contribuintes do Alto Rio das Mortes (**Figura 1**), totalizando 109 coletas efetuadas entre Janeiro de 2002 e março de 2003. As análises de pesticidas foram efetuadas junto ao Departamento de Química da UFMT, utilizando um Cromatografo a gás Agilent, modelo HP 6890 com detector de massa modelo HP 5973. Para avaliação por regressão logística foi efetuada, para cada coleta,

contagem da presença/ausência das principais 18 pesticidas utilizadas na região. Para binarização foram utilizados valores de corte 1 e 2 (1 (2) ou mais pesticidas presentes na amostra).

Os modelos de risco para presença dos pesticidas foram desenvolvidos a partir de Regressões Logísticas (RL) (Hosmer & Lemeshow 1989). RL aplicam-se na modelação de dados dependentes de natureza dicotômica ou binária. A RL estima a probabilidade da ocorrência de um evento e pode ser definida, de forma geral, como:

$$g(P_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad [1]$$

onde:

$g(P_i)$ = função de ligação

P_i = probabilidade de resposta para o i -ésimo fator (ou co-variante)

β_i = constantes

x_i = variáveis independentes

Na função de ligação do tipo *logit* com distribuição logística, $g(P_i)$ é dado por $\ln(P_i)/(1 - P_i)$. É chamada “*odds ratio*” a razão $P_i / 1 - P_i$ que pode ser interpretada como probabilidade de um evento ocorrer. Para avaliação da confiabilidade dos modelos, utilizamos a estimativa do nível de significância dos coeficientes determinados. A tabela de classificação, na qual se obtém a taxa dos casos que são classificados corretamente pelo modelo e o coeficiente de Cox & Snell, compara as probabilidades do modelo com as de um modelo reduzido baseado somente na constante determinada.

4. Resultados e Discussões

A **figura 2** mostra os seis PIs temáticos elaborados para seu consecutivo cruzamento com os resultados de monitoramento de qualidade de água nas exutórias das sub-bacias amostradas.

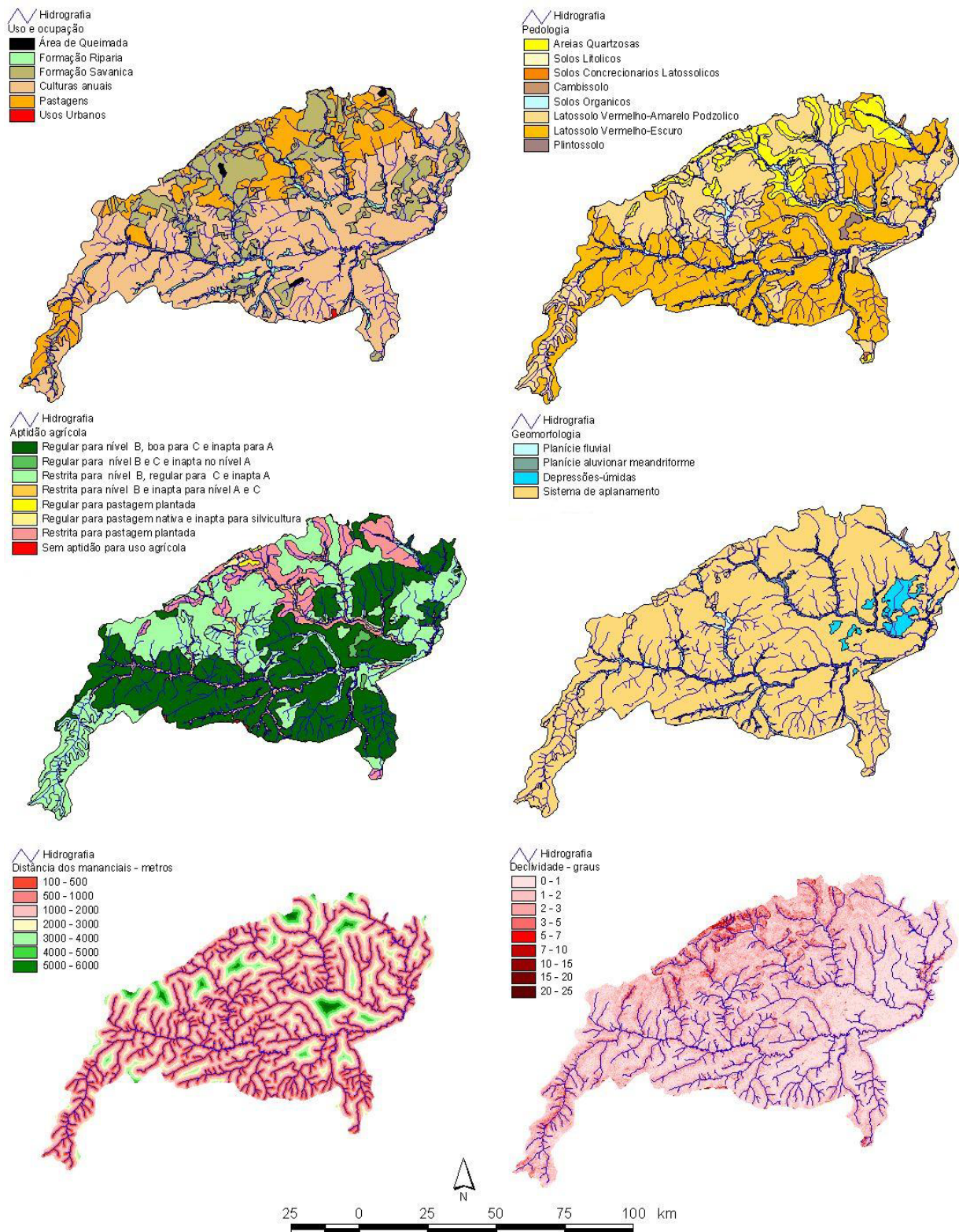


Figura 2. Planos de informação avaliados para desenvolvimento dos modelos de risco de contaminação de águas superficiais por agrotóxicos.

Além de uma variável não-espaciaizada “Período sazonal”, foram avaliadas as supostas relações com os PIs do “Uso e ocupação de solo”, das “unidades pedológicas”, “aptidão agrícola”, “unidades geomorfológicas”, “declividade” e “distância para o manancial mais

próximo”. A variável “distância para o manancial mais próximo” foi logaritimizada para homogeneização dos dados.

Para introdução na RL, as categorias de três PIs com escala nominal (Pedologia, Aptidão Agrícola, Geomorfologia) foram re-codificadas em escala ordinal, atribuindo valores crescentes de acordo com a suposta intensidade dos impactos negativos da categoria na qualidade de água. Em seguida foram calculadas as medianas ponderadas por bacia de contribuição, de acordo com os valores ordinais. Do PI de “Uso e Ocupação de Solo” foi determinada a porcentagem da área total ocupada por lavouras. A partir dos PIs da Distância dos mananciais e Declividade foram calculadas as respectivas médias aritméticas por sub-bacia. Análises de regressão simples mostraram forte colinearidade para as variáveis independentes “Geomorfologia”, “Pedologia” e “Aptidão agrícola”. Em consequência, somente a variável “Aptidão agrícola” foi mantida para cálculo do modelo final de RL, variável cujas classes são parcialmente determinadas por aspectos da “Geomorfologia” e “Pedologia”. A **tabela 2** mostra as principais saídas do modelo de RL (*backward conditional*), que indicou o período sazonal (“Chuva”) e a “Distância para os mananciais” (“Distancia”) como fatores explanatórios significativos. A tabulação cruzada na **tabela 3** mostra desempenho satisfatório do modelo (acerto de 86,2%).

Tabela 2. Principais indicadores estatísticos do modelo de RL (*backward conditional*) para mapeamento de risco de contaminação por pesticidas.

		B	S.E.	Wald	df	Sig.	Exp(B)
1ª Etapa ^a	Chuva	6,952	2,175	10,216	1	,001	1045,055
	Declive	12,157	9,661	1,584	1	,208	190406,588
	Distancia	-27,695	19,003	2,124	1	,145	,000
	UsoSolo	20,006	15,249	1,721	1	,190	488029736,69
	Aptidao	5,712	4,780	1,428	1	,232	302,596
2ª Etapa	Chuva	7,213	2,615	7,609	1	,006	1357,213
	Declive	,648	1,249	,269	1	,604	1,912
	Distancia	-5,525	2,163	6,525	1	,011	,004
	UsoSolo	2,760	3,203	,743	1	,389	15,804
3ª Etapa	Chuva	7,462	2,870	6,762	1	,009	1741,343
	Distancia	-5,365	2,282	5,527	1	,019	,005
	Usosolo	2,591	3,131	,684	1	,408	13,338
4ª Etapa	Chuva	7,101	2,521	7,935	1	,005	1212,913
	Distancia	-4,379	1,675	6,837	1	,009	,013

^aVariáveis inclusas na 1ª Etapa: Chuva, Declive, Dist_Log, UsoSolo, Aptidão.

Tabela 3: Tabulação cruzada entre os valores reais observados e os valores previstos pela RL.

Observado	Previsto		
	CUTO		Porcentagem de classificação correta
	0	1	
4ª Etapa CUTO	48	15	76,2
1	0	46	100,0
Porcentagem Total			86,2

A figura 3 representa a simulação espacial da probabilidade da presença de pesticidas nas sub-bacias do Alto Rio das Mortes. Para representar o valor 2, que significa o período de chuvas da variável “Período sazonal” na tabela de atributos, criou-se um PI que representava a espacialização desta variável.

Para a geração do modelo, fez-se o cruzamento dos PIs “Distancia para o manancial” e “Período Sazonal”, aplicando os coeficientes obtidos pelo modelo de Regressão Logística. As partes setentrionais da bacia foram excluídas do modelo, considerando que sua ocupação por áreas de plantio é marginal, enquanto as sub-bacias amostradas pelo monitoramento de qualidade de água se encontram em regiões com uma ocupação de no mínimo 70% por plantio intensivo.

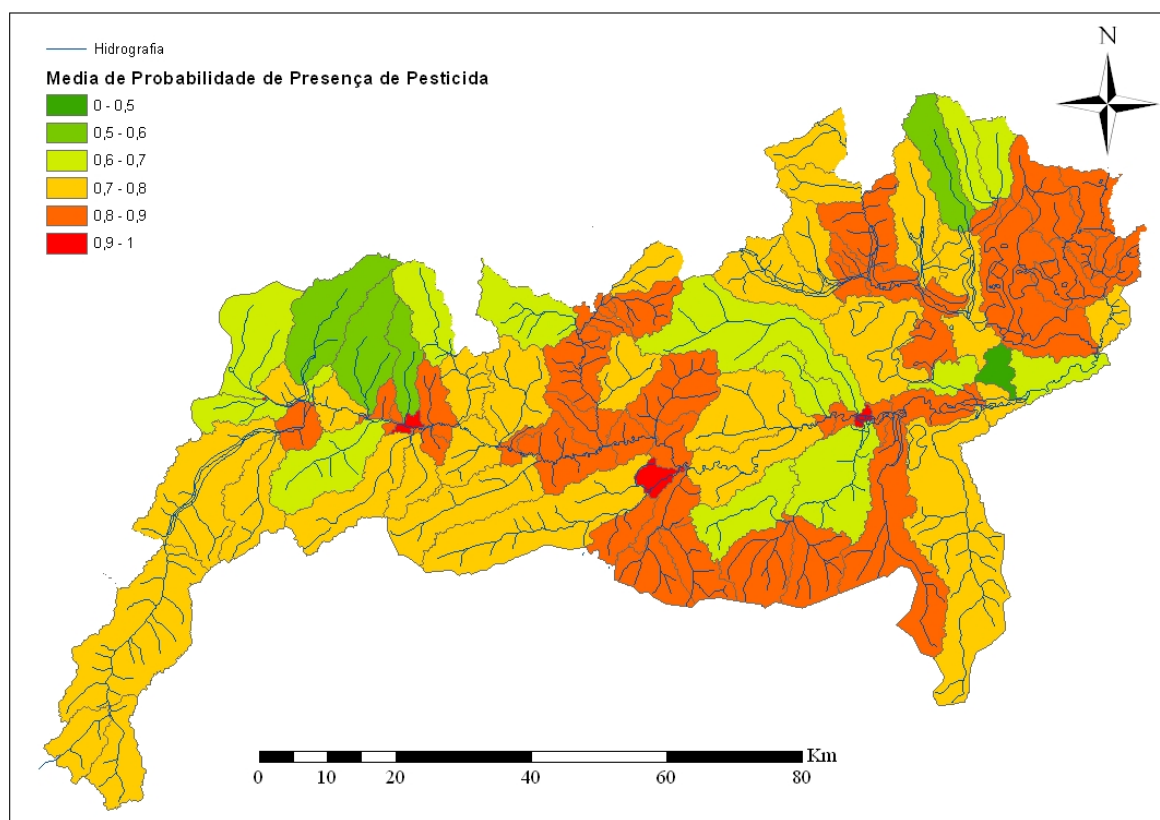


Figura 3: Mapa de distribuição da probabilidade média de presença de pesticidas.

5. Conclusões

Os resultados pré-liminares apresentados neste estudo permitem as seguintes conclusões:

- Técnicas de geoprocessamento constituem, na área de estudo, em uma ferramenta valiosa para parametrização de bacias de contribuição, visando uma avaliação das suas interferências na presença de agrotóxicos nas mananciais superficiais;
- O modelo de RL, entretanto, mostrou somente poder preditivo significativo para uma variável espacializada, a da “Distância para o manancial”, indicando a grande importância da proteção das mata ciliares para a proteção áreas de mata ciliares
- Supõe-se que a não-consideração de fatores como “Uso e ocupação de solo” no modelo se deu devido a característica das sub-bacias amostradas: todas possuem uma ocupação de no mínimo de 70% da sua área com labores de plantio intensivo
- Para o desenvolvimento de um modelo igualmente válido para a bacia toda devem ser inclusos nas análises de qualidade de água bacias de uso misto e maior índice de conservação da vegetação nativa;

6. Referências

- Battaglin, W. A.; Goolsby, D.A. Regression models of herbicide concentrations in outflow from reservoirs in the midwestern USA, 1992–1993. **Journal of the American Water Resources Association**, vol. 34, no. 6, p. 1369–1390, 1998.
- Battaglin, W.A., Ulery, R.L., Winterstein, T., and Welborn, T., 2003, **Estimating the Susceptibility of Surface Water in Texas to Nonpoint-Source Contamination by Use of Logistic Regression Modeling**. Washington, DC: U.S. Geological Survey Water-Resources Investigations Report 03-4205, 24 p.
- Bengraïne, K.; Marhaba, T. F. Using principal component analysis to monitor spatial and temporal changes in water quality. **Journal of Hazardous Materials B**, vol. 100, no. 1-3, p. 179–195, 2003.
- Dores, E. F. G.C.; De-Lamonica-Freire, E. M.. Aquatic environment contamination by pesticides. Case study: water used for human consumption in Primavera do Leste, Mato Grosso - preliminary analyses. **Química Nova**, v. 24, n. 1, 27-36, 2001.
- Donohue, I.; Styles, D.; Coxon, C.; Irvine, K. Importance of spatial and temporal patterns for assessment of risk of diffuse nutrient emissions to surface waters. **Journal of Hydrology** , vol. 304, p. 183–192, 2005.
- Greene, E. A.; LaMotte, A. E.; Cullinan, K.-A., 2004. **Ground-Water Vulnerability to Nitrate Contamination at Multiple Thresholds in the Mid-Atlantic Region Using Spatial Probability Models**. Denver, CO: U.S. Geological Survey, Scientific Investigation Report 2004-5118, 32 p.
- Munafò M.; Cecchi, G.; Baiocco, F.; Mancini, L. River pollution from non-point sources: a new simplified method of assessment. **Journal of Environmental Management**, v. 77, p. 93–98, 2005.
- Wunderlin, D. A.; Diaz, M. P.; Amé, M. V.; Pesce, S. F.; Hued, A. C., Bistoni, M. L. A. Pattern Recognition Techniques for the Evaluations of Spatial and Temporal Variations in Water Quality, a Case Study: Suquia River Basin (Córdoba Argentina). **Water Research**, v. 35, n. 12, p. 2881-2894, 2001.